



## Online short-term forecast of greenhouse heat load using a weather forecast service

Vogler-Finck, P. J.C.; Bacher, P.; Madsen, Henrik

*Published in:*  
Applied Energy

*Link to article, DOI:*  
[10.1016/j.apenergy.2017.08.013](https://doi.org/10.1016/j.apenergy.2017.08.013)

*Publication date:*  
2017

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Vogler-Finck, P. J. C., Bacher, P., & Madsen, H. (2017). Online short-term forecast of greenhouse heat load using a weather forecast service. *Applied Energy*, 205, 1298-1310.  
<https://doi.org/10.1016/j.apenergy.2017.08.013>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Online short-term forecast of greenhouse heat load using a weather forecast service

P.J.C. Vogler–Finck<sup>a,b,\*</sup>, P. Bacher<sup>c</sup>, H. Madsen<sup>c</sup>

<sup>a</sup>*Neogrid Technologies ApS, Niels Jernes Vej 10, 9220 Aalborg Ø, Denmark*

<sup>b</sup>*Department of Electronic Systems at Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg Ø, Denmark*

<sup>c</sup>*Department of Applied Mathematics and Computer Science at the Technical University of Denmark, Artusvej 5, 2800 Kgs. Lyngby, Denmark*

---

## Abstract

In some district heating systems, greenhouses represent a significant share of the total load, and can lead to operational challenges. Short term load forecast of such consumers has a strong potential to contribute to the improvement of the overall system efficiency. This work investigates the performance of recursive least squares for predicting the heat load of individual greenhouses in an online manner. Predictor inputs (weekly curves terms and weather forecast inputs) are selected in an automated manner using a forward selection approach. Historical load measurements from 5 Danish greenhouses with different operational characteristics were used, together with weather measurements and a weather forecast service. It was found that these predictors of reduced complexity and computational load performed well at capturing recurring load profiles, but not fast frequency random changes. Overall, the root mean square error of the prediction was within 8–20% of the peak load for the set of consumers over the 8 months period considered.

**Keywords:** heat demand, load forecast, recursive least squares, weather forecast service, greenhouses, model selection

---

## 1. Introduction

### 1.1. Motivation and problem statement

In northern European countries, the growing of certain crops requires the use of heated greenhouses. For example, on the Danish island Funen, greenhouses are mostly connected to the district heating, where they are among the most sensitive loads of the

---

\*Corresponding author: pvf@neogrid.dk

system and represent a significant share (close to a quarter) of the heat load in the system.

To improve the efficiency of the district heating system operation and lower the losses, forward temperatures should be kept as low as possible while still meeting the needs of the consumers. A good medium and short term forecast of heat demand is therefore of key importance in improving planning and operation of the system [1].

In this study, the heat load of five greenhouses is predicted in an online adaptive manner using the recursive least-squares (RLS) method. The scope is limited to *short-term* forecast, which here refers to predictions from one hour to less than one week ahead (as opposed to *long-term* forecast used for system planning). Historical heat load data from a Danish utility is used, as well as weather data from both local measurements and forecasts from a third party. Moreover, an automated explanatory variable selection algorithm is introduced to reduce the need for man-hours in individual load modelling. The simplicity of the approach (combining reduced complexity, low computational burden, and automated modelling process) makes it a well-suited candidate for widespread practical applications, on the condition that it achieves sufficient performance – which is evaluated in this work.

## 1.2. State of the art

Load forecast in energy systems using weather parameters is not a new problem, as it has been studied for decades both in power systems and district heating systems [2, 3]. However, the case of power systems seems to have been covered more extensively than district heating systems [4]. In this work, the focus is on heat loads within district heating systems (a reader interested in more details on power load prediction may refer to reviews [5, 6] for further reading).

Systems for load forecasting are an important lever for obtaining large savings in district heating systems. In particular, short term predictions of heat load are required in advanced controllers in district heating networks (see [7, 8, 9] for predictive/optimal control, and [10] for multi-agent based control), as well as forecasts of water temperatures at given points in a network which are expected to significantly reduce the heat production and supply costs [11].

However, very limited literature seems to be available on greenhouses heat load forecast. Daily heat loads were predicted in [12] and [13] in an offline manner using knowledge of the indoor temperature of the greenhouses. Energy demands of greenhouses were predicted with an hourly resolution in [14] (where the energy demand was limited to heating and extensive knowledge of the greenhouse was available in the modelling phase) and [15] (where the energy demand included both heat and equipment) using knowledge of internal parameters (humidity and temperature). To the best of our knowledge, online short-term forecast of heat loads with sub-daily resolution without knowledge of internal parameters of the greenhouses has not been covered in previous work (less so within district heating systems).

Various methods and tools from statistics and computer science have been applied to tackle the problem of heat load forecasting for different systems, including among others: recursive least squares using simple linear models [16, 17, 18], multiple linear regression [19], autoregressive models with exogenous inputs [9, 20], seasonal autoregressive integrated moving average [21], classification and regression tree [19], neural networks [19, 22] (multiple neural networks are proposed in [23] in the different context of gas networks, a comparison of radial basis function network feedforward backpropagation networks, neuro-fuzzy interference systems and their combination is found in [24], and a more detailed review for the case of power demand is made in [25]) and neuro-fuzzy logic [26] (an adaptive neuro-fuzzy logic inference system is presented in [27] in the context of electricity demand), support vector machines [19] (combined with a discrete wavelet transform algorithm [22], or with either polynomial or radial basis function as kernel functions [28]), combination of wavelet-based multi-resolution analysis and artificial neural networks [8, 29], and genetic programming [22]. K-nearest neighbour and Markov-chains are additional possibilities of prediction algorithms [30]. Combination of the output of individual forecasting methods can also result in performance improvement [24, 31]. For further insight, one may refer to the review [4, Chap. 5] for energy demand forecasts with intraday resolution, and the review [32] for a wide spectrum review of heating (as well as cooling and power) demand in buildings ranging from short-term online forecasting to offline long term forecasting.

Typical external input variables to the predictors are ambient (outdoor) temperature [16, 18, 20, 22, 28, 33] (both daily mean and maximum were used in [24] where a daily resolution was adopted), supply temperature of the water in the district heating system [22], global sun radiation [16, 18, 20, 24], and wind speed [16, 18, 20]. Humidity was also mentioned as a potential explanatory variable [19, 20, 24], although [20] discarded it as it was not statistically significant in the chosen model. Lagged versions of all of those inputs were also found to be relevant, and used either explicitly [22, 33] or implicitly through the use of filters on the input signals [16, 18]. In order to account for time-dependencies in weather uncertainty, ensemble forecast were also investigated for heat load forecast in [20] and for power load forecast in [34].

Additionally, introduction of time dependency and periodicities can be achieved through the use of diurnal curves (with the possibility of using separate curves for weekends and weekdays [16]), explicit use of day of the week and month of the year [24], or a binary workday/weekend input [9].

On an overall system level and based upon conclusions from [35], [36] states that one of the greatest influences on the heat load are from ambient temperature and social behaviour of the consumers (i.e. time dependency), while wind, global radiation and precipitation have a lower importance. A separation of the load in a sum of an ambient temperature dependent component (modelled through piecewise linear functions) and a day profile is then proposed.

Internal states of the modelled heat load as explanatory variables in some studies, such as: return temperature [22, 33] and flow rates [19] of the water in the district heating system, as well as heat load at previous times [9, 20, 22, 24]. In the specific case of greenhouses, additional relevant internal parameters are indoor relative humidity and air temperature [14] and radiative losses to the sky [13].

Identification of relevant explanatory variables (also known as features) from a set of available variables is a classic problem in statistical learning known as the *subset selection problem*, which has been studied for decades [37]. Two types of relevance were proposed in [37]: *strong* relevance where the removal of the feature *always* decreases accuracy, and *weak* relevance where this removal *sometimes* leads to loss of accuracy. Such a selection of explanatory variables of a model is typically made manually in a *forward selection* manner where new variables are added one by one. Typical selection criterion used in this process are  $R^2$  increase [24], root mean square error (RMSE) decrease [20], Akaike Information Criterion (AIC) decrease [20] or cross-correlation between residuals and inputs [16].

### 1.3. Contribution

Although the recursive least squares framework is well established in the literature, an automated selection of input signals based upon root mean square error criterion was not encountered by the authors in previous applications from the literature. The associated algorithm is the first contribution of this paper and a step towards automated model fitting for this application.

A second contribution is in the experimental quantification of the performance of this short-term heat load forecast method on real greenhouses hourly data. In particular, a combination of an online adaptive approach and the use of a commercially available weather forecast service were not found in prior works (traditionally, most works in this field use offline approaches using weather measurements only).

A further contribution lies in the evidence that different input signals are relevant for different greenhouses (including frequencies lower than daily, which were not found in previous literature).

## 2. Methodology

In this part, the data collection and modelling framework are presented.

### 2.1. Data collection

#### 2.1.1. Heat load

Heat load data from a number of greenhouses was provided by *Fjernvarme Fyn A/S*, the district heating operator of the Danish island Funen. Among all those greenhouses, five of them were selected to convey a more detailed analysis. Attention was given to select greenhouses with different characteristics in order to have a varied sample. The consumption data of those 5 greenhouses was sampled every 15 min (which was re-sampled to hourly values for this study), and spans over an 8 months period (June 2015 to the end of January 2016), as presented in Fig. 1.

As seen on this figure, these greenhouses use heating in very different ways, although most of their consumption is happening in spikes rather than a smooth profile. Summary statistics of the heat load for each of them are given in Tbl. 1, and normalised

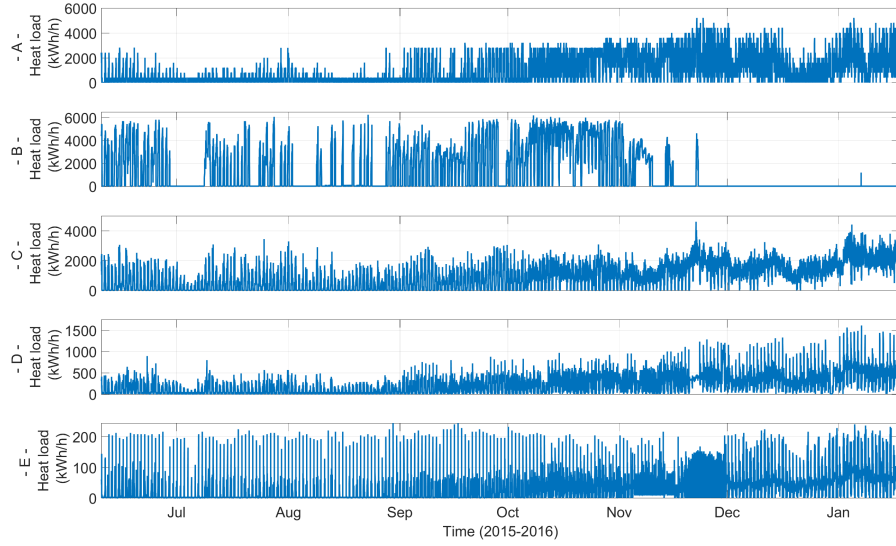


Figure 1: Heat load data for the selected greenhouses (15 min sample time, prior to re-sampling to hourly values).

Table 1: Summary statistics of the load data (units: MW).

| Greenhouse | Min | Max | Mean | Std  |
|------------|-----|-----|------|------|
| A          | 0   | 5.2 | 0.9  | 1.1  |
| B          | 0   | 6.2 | 1.2  | 1.8  |
| C          | 0   | 4.6 | 0.9  | 0.8  |
| D          | 0   | 1.6 | 0.2  | 0.2  |
| E          | 0   | 0.2 | 0.03 | 0.04 |

*load-duration curves* [38, 39] are presented in Fig. 2 (normalisation is made using the measured peak load, which is expected to be close to the sizing of the connection to the district heating system).

Greenhouse A uses little heat in summer, and increases its load as winter settles (a reduction of heat load is observed in the second half of December). Greenhouse B exhibits periods with no heating use, including a total shut-down of the heating in December and January. Greenhouses C and D have similar behaviours, with low load in summer and an increase in winter. Greenhouse E shows load spikes in both winter and summer, with a higher load in the winter-time.

In practise, the district heating operator rarely obtains any feedback from the owners of the greenhouses (apart from incidents or major load management events). Therefore, it is assumed that the only available information about the greenhouse states are the readings from the local heat meter. It is however worth remembering that knowledge of internal states of the greenhouses (such as indoor humidity and air temperature) may provide further improvement.

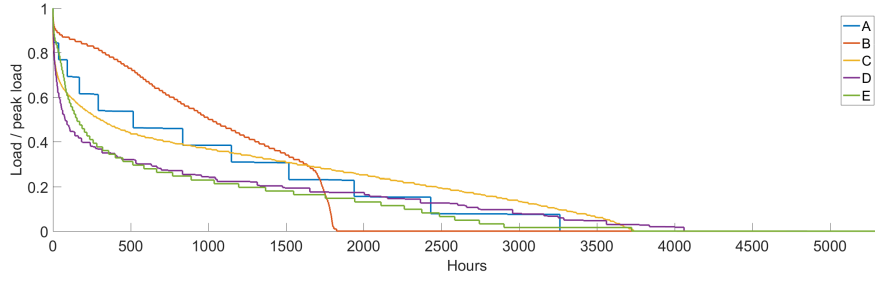


Figure 2: Normalised load-duration curve (using 15 min sample time, prior to hourly resampling).

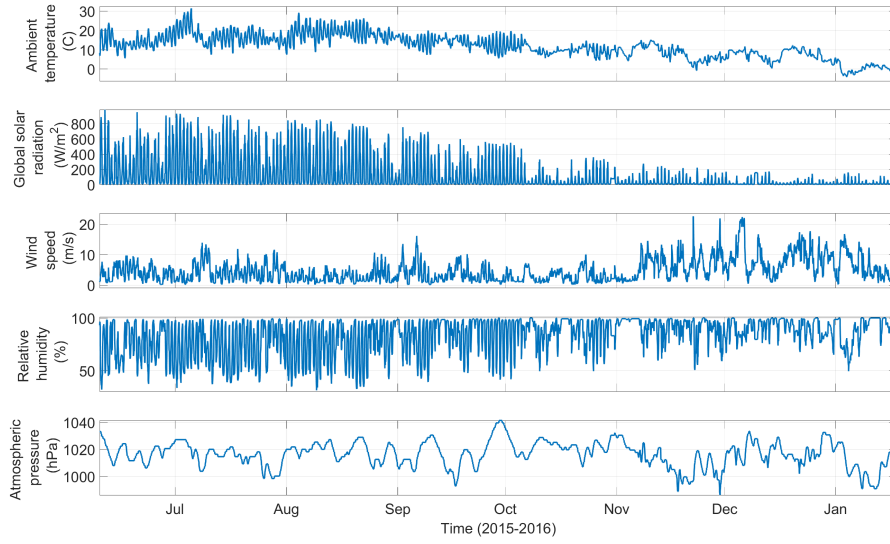


Figure 3: Measurements from the district heating operator weather station.

### 2.1.2. Weather measurements

The district heating operator has its own weather station in a neighbouring city, located within 20 km of the greenhouses location. Those weather observations from the corresponding period were also provided by the district heating system operator.

The following parameters were measured with an hourly resolution: ambient temperature ( $T_a$  in  $^{\circ}\text{C}$ ), global sun radiation ( $\Phi_G$  in  $\text{W}/\text{m}^2$ , average over the last hour), wind speed ( $W_s$  at 10 m mast height), relative humidity ( $h$  in %), and atmospheric pressure ( $p_A$  in hPa). Corresponding timeseries are plotted in Fig. 3 and summary statistics of the data given in Tbl. 2.

Table 2: Statistical properties of the weather data.

| Parameter                    | Min  | Max  | Mean   | Std   |
|------------------------------|------|------|--------|-------|
| $T_a$ (C)                    | -7.6 | 31.4 | 9.0    | 6.2   |
| $\Phi_G$ (W/m <sup>2</sup> ) | 0    | 1011 | 124.4  | 204.0 |
| $W_s$ (m/s)                  | 0.07 | 22.5 | 5.4    | 3.8   |
| $h$ (%)                      | 23.5 | 100  | 81.4   | 17.4  |
| $p_A$ (hPa)                  | 976  | 1041 | 1015.5 | 10.6  |

Table 3: Quantification of forecast error.

| Parameter                    | MAE range | RMSE range | Bias range     |
|------------------------------|-----------|------------|----------------|
| $T_a$ (°C)                   | 0.7–1.1   | 0.9–1.5    | −0.004 – 0.035 |
| $\Phi_G$ (W/m <sup>2</sup> ) | 29–35     | 59–67      | 0.87 – 2.60    |
| $W_s$ (m/s)                  | 1.8–2.0   | 2.3–2.5    | 0.29 – 0.96    |
| $h$ (%)                      | 8.1–9.2   | 9.7–11.1   | −5.3 – −4.2    |

### 2.1.3. Weather forecast

A weather forecast service, provided by *ENFOR A/S*, was used for the location over the period of the study. This forecast provides hourly estimates of ambient temperature (at 2 m height), global sun radiation (average over the last hour), wind speed (at 10 m mast height) and relative humidity over a prediction horizon of 6 days. Calibration of the forecast of the ambient temperature was implemented using local measurements. This forecast is updated every hour.

Although this study restricts the prediction to the next 48 h, in practise the limiting factor for the heat load prediction horizon would be the availability of weather forecast data (here 6 days ahead), the computational power at hand, online latency of readings, and the precision achieved by the model.

An uncertainty analysis of the forecast over its horizon was conveyed, with the results presented in Fig. 4. Prediction errors (in the mean average error (MAE) and root mean square error (RMSE) sense) for the range of horizons considered are detailed in Tbl. 3. It is worth knowing that the negligible bias on  $T_a$  originates from regular calibration of ambient temperature against historical measurements. A significant amount of outliers was observed, which suggests that the distribution of residuals would not be well represented by a Gaussian distribution.

## 2.2. Data preparation

First of all, the data from the different parties were assembled into synchronised hourly measurements. In rare cases, some measurements values were either missing or corrupted and replaced by the previous valid measurements was made. This option was preferred over ignoring or zeroing the corresponding values, in order to limit the risk of resulting perturbations of the adaptive model.



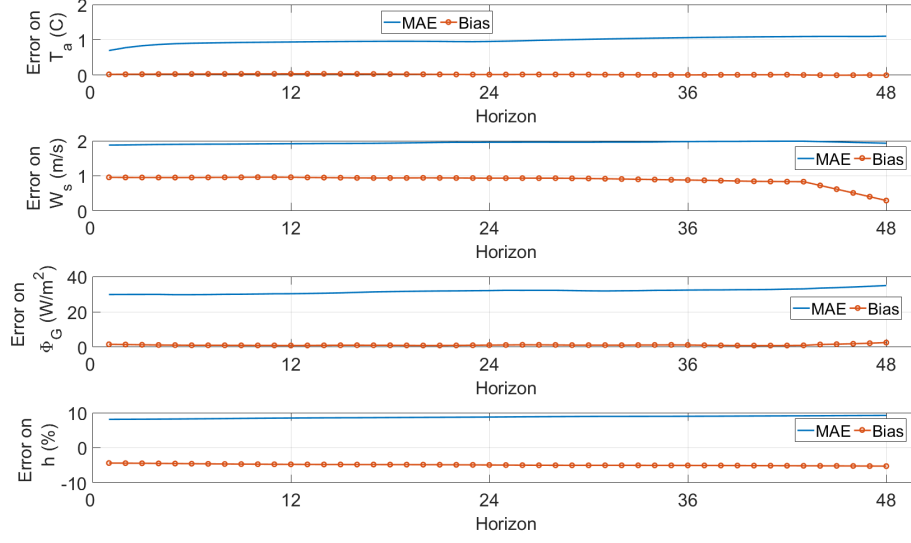


Figure 4: Forecast error with horizon on the different predicted variables (mean average error (MAE) and bias).

For greenhouses A, C, D and E, the whole dataset from June 10th 2015 to January 18th 2016 was considered. For Greenhouse B, the study was limited to the period from June 10th 2015 to November 26th 2015 which is the period in which heating was used in the building, as seen in Fig. 1.

### 2.3. Modelling of time dependency

In order to model time dependency, additional inputs timeseries are generated in the form of weekly curves, similarly to [16]. Such weekly curves express periodicity of the behaviour. Formally, this translates into the Fourier series:

$$\begin{aligned} C_i[j] &= \cos\left(ij \frac{2\pi T_s}{T_0}\right) \\ S_i[j] &= \sin\left(ij \frac{2\pi T_s}{T_0}\right) \end{aligned} \quad (1)$$

where  $T_0 = 1$  week is the period of the fundamental,  $T_s = 1$  h is the sample time,  $j$  is the time of the sample considered (discrete).

In order to avoid aliasing of the weekly curve terms with the hourly sampling, only the first 83 harmonics were considered in the study. However, as will be seen later as a result of input selection, only a reduced number of these were found to be relevant inputs for the model.

#### 2.4. Candidate model input parameters

The input candidates for the model are the following:

- Weekly curve terms ( $C_i, S_i$ ) for  $i \in \{0, 1, \dots, N_{wc}\}$  ( $N_{wc} = 84$ , the 0 term corresponds to constant 1 value)
- Ambient temperature  $T_a$
- Global solar radiation  $\Phi_G$
- Wind speed  $W_s$
- Relative humidity  $h$
- Atmospheric pressure  $p_A$

To account for dynamics in the weather dependency, a simple first order filter was used on weather inputs ( $T_a, \Phi_G, W_s, h, p_A$ ), similarly to [16]:

$$H_a(q) = \frac{1 - a}{1 - a q^{-1}} \quad (2)$$

Where  $q^{-1}$  is the backward shift operator (i.e.  $q^{-1} z[j] = z[j - 1]$ ), and  $a \in [0, 1]$  is the filter parameter to be fitted separately for each input high inertia tends to lead to a high value of  $a$ , whereas low inertia leads to lower values).

#### 2.5. Recursive least-squares

As some of the greenhouses have an activity that varies throughout the year (e.g. due to changes to another crop type requiring different conditions – see [13] for an example of differing temperature needs), time-adaptive models were sought in order to improve the robustness of the forecast. Since the project aimed at handing over some of the methods to the district heating operator as part of the collaboration, a simple approach was sought. Thus, linear models with recursive least-squares implementation are preferred here, due to their simplicity, ease of tuning and low computational cost compared to the more complex alternatives applied in other studies.

An in-depth presentation of the recursive least-squares framework is presented in [40, Chapter 11] and the initial report of the study [41]. A forgetting factor was used in order to allow for model adaptation to changes in greenhouse operation over time (a value of 0.99 was chosen here). As heat loads of greenhouses were always positive (none of them had local heat production for the district heating), an additional adjustment was made to the classic recursive least squares framework by imposing a minimum value of 0 as an output of the predictor.

In this study, a period of two weeks was used to initialise the model. This model was then recursively updated every hour using the measured heat load. It is worth noting that there is no need for a separation of data into a training and validation set in the recursive least squares approach. The model updates itself continuously over time using only the past data for learning, and predicts the future load.

### 2.6. Input selection

In order to select appropriate inputs, a forward selection of inputs was adopted. An automated approach, relying upon the reduction of the root mean square error (RMSE over the prediction period, as defined in Eq. 3) is used, as it gives an optimal estimate of the conditional mean expectation (conditioned on the inputs) [40].

$$\text{RMSE}(\epsilon) = \sqrt{\frac{1}{N} \sum_{k=1}^N \epsilon[i]^2} \quad (3)$$

### 2.7. Setups for prediction

In order to evaluate the improvement due to the addition of a weather forecast, 3 setups were considered:

**No forecast** where no knowledge of future weather inputs is available. In this case, the present values of those inputs variables are used.

**Causal forecast** where future values of the weather inputs are taken from the weather forecast.

**Perfect forecast** where future values of the weather inputs are taken from the actual measurements.

A visual summary of the forecasting system is provided in Fig. 5.

## 3. Results

In this section, the results of the study are presented.

### 3.1. Analysis of relevant parameters

The forward selection procedure was applied on all 5 greenhouses to identify relevant explanatory variables, using data covering the whole period. Relevant inputs for the forecast based upon weather forecast inputs are presented in Tbl. 4 and improvement along the forward selection process is shown in Fig. 6. It appears that the variables leading to the strongest reduction in RMSE (constant term excluded) are the 7th term of the weekly curve and ambient temperature (noticeably more than every other relevant weather parameter).

In this study, heat load at previous times was not used as an explanatory variable. However, the model update in the recursive least squares approach implicitly accounts for past behaviour in the resulting model.

Relevant weather variables were occasionally found to differ between setups. For Greenhouse A,  $\Phi_G$  was not selected as a relevant input in the *no forecast* setup. For

**Result:** Vector of relevant explanatory variables  $X$  and best model performance

```

    RMSEbest
 $X \leftarrow ()$ ;
    RMSEbest  $\leftarrow +\infty$ ;
    for  $i$  from 0 to  $N_{wc} + 5$  do
        switch  $i$  do
            case  $i=0$  do
                |  $X_{add} \leftarrow (\mathbb{1}[k])$ 
            case  $1 \leq i \leq N_{wc}$  do
                |  $X_{add} \leftarrow \begin{pmatrix} C_i[k] \\ S_i[k] \end{pmatrix}$ 
            case  $i = N_{wc} + 1$  do
                |  $X_{add} \leftarrow (H_{T_a}(q)T_a)$ 
            case  $i = N_{wc} + 2$  do
                |  $X_{add} \leftarrow (H_{\Phi_G}(q)\Phi_G)$ 
            case  $i = N_{wc} + 3$  do
                |  $X_{add} \leftarrow (H_{W_s}(q)W_s)$ 
            case  $i = N_{wc} + 4$  do
                |  $X_{add} \leftarrow (H_h(q)h)$ 
            case  $i = N_{wc} + 5$  do
                |  $X_{add} \leftarrow (H_{p_A}(q)p_A)$ 
            end
        end
         $X_{fs} \leftarrow \begin{pmatrix} X \\ X_{add} \end{pmatrix}$ ;
        compute prediction  $\hat{Y}$  over the whole dataset using  $X_{fs}$  as input vectors ;
        if problem is numerically ill conditioned then
            | terminate current loop iteration and continue to next step  $i$ 
        end
        compute residuals  $\epsilon$  ;
        RMSE $i$   $\leftarrow$  RMSE( $\epsilon$ ) ;
        if RMSE $i$  < RMSEbest then
            |  $X \leftarrow X_{fs}$ ;
            | RMSEbest  $\leftarrow$  RMSE $i$ ;
        end
    end
end

```

**Algorithm 1:** Forward selection algorithm.

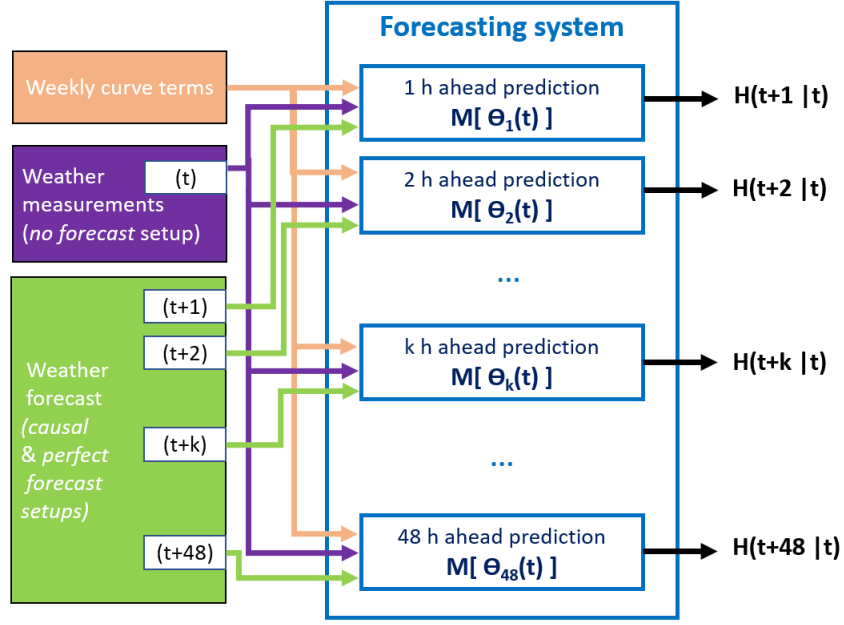


Figure 5: Structure of the forecast. The common model structure  $M$  is identified *a posteriori* on the 1 h ahead prediction, and parameter vectors  $\theta_1, \dots, \theta_{48}$  are identified (and updated) separately for each number of steps ahead using the RLS approach.

Table 4: Selected relevant inputs per greenhouse (setup *causal forecast* - weekly curve terms multiple of 7 express daily periodicity) .

|   | Weekly curve terms                                 | Weather inputs        |
|---|--|-----------------------|
| A | 0, 1, 6, 7, 14, 21, 28, 35, 49, 56                 | $T_a, \Phi_G, W_s, h$ |
| B | 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 13, 14, 21           | $T_a, \Phi_G, W_s, h$ |
| C | 0, 1, 6, 7, 8, 13, 14, 21, 28, 35, 42, 56, 77      | $T_a, \Phi_G, W_s, h$ |
| D | 0, 1, 6, 7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77 | $T_a, W_s$            |
| E | 0, 7, 14, 21, 28, 35, 42, 49, 56, 63, 70, 77       | $T_a$                 |

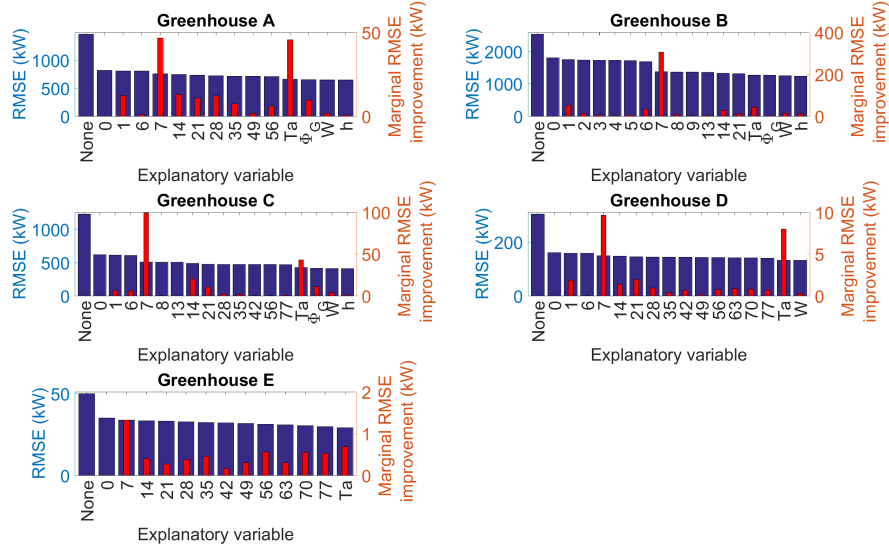


Figure 6: Evolution of the RMSE along the forward selection process (the initial value for no explanatory variable corresponds to the standard deviation of the original load signal, numbers refer to the corresponding weekly curve term). The value of the RMSE is shown in blue, and marginal improvement of RMSE for the addition of each variable is shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Greenhouse D,  $h$  was not selected in the *causal forecast* setup,  $\Phi_G$  was selected in the *perfect forecast* setup. For Greenhouse E,  $h$  selected in the *no forecast* and *perfect forecast* setups, and  $\Phi_G$  was selected in the *perfect forecast* setup.

As Tbl. 4 shows, the relevant inputs differ between greenhouses, both for weekly curve terms and weather variables. This supports the use of a forward selection approach. Constant term,  $7^{th}$ ,  $14^{th}$  and  $21^{st}$  terms of weekly curve (i.e. period of 1 day, 12 h and 8 h, indicating a clear daily periodicity) and ambient temperature were always found to be relevant here. Wind speed was identified relevant in 4 cases, global solar radiation and humidity were found relevant for 3 of the greenhouses. On the other hand, atmospheric pressure was never found to be a relevant explanatory variable on this sample. The results also suggests that larger consumers (greenhouses A,B,C,D) may be more affected by all weather inputs and have slower dynamics than smaller ones (Greenhouse E); the sample size of the study does however not allow drawing any conclusion on this.

Residuals of the 1 h ahead predictor are presented in Fig. 7. In each case, a correlation of the residuals 24 h later is observed. Autocorrelation in the residual were observed to be low for greenhouses A, C, D, and E, and strong for Greenhouse B, suggesting the need for error modelling, which has been investigated below.

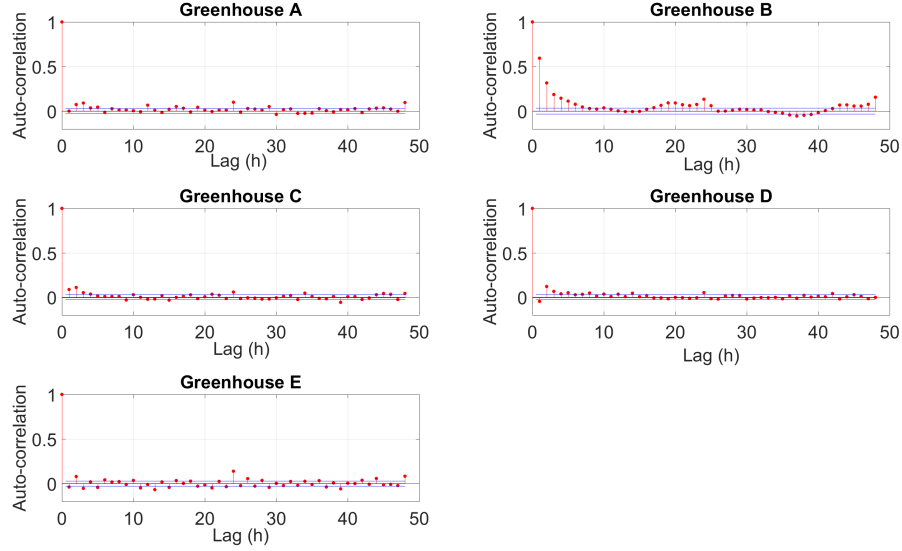


Figure 7: Autocorrelation of the 1 hour ahead forecast error (*causal forecast setup*).

### 3.2. Benefits from a weather forecast

Comparison of the forecast performance in the 3 setups (with *no*, *causal* and *perfect weather forecast*) is presented in Fig. 8. Apart from Greenhouse B, the weather forecast was always found to improve the performance of the predictor, with a greater benefit for longer horizon as expected.

In every case, *perfect forecast* resulted in a constant RMSE over the 48 h horizon. This may be explained by the slowness of changes in the model due to the high value of the forgetting factor used here.

Forecast using *causal* and *perfect weather forecast* provided comparable performances. In the case of Greenhouse B, performance was even improved, which may be a sign of over-fitting resulting from collinearity between weather forecast and weekly curves.

Compared to the *no forecast* setup, the *causal forecast* reduced the RMSE by 2.8–6.4% for Greenhouse A, 0.6–3.5% for B, 3.2–9.2% for C, 1.0–4.3% for D, and 0.3–2.0% for E.

From this point on, only the *causal weather forecast* setup is considered.

### 3.3. Performance of predictor using causal forecast

In all 5 cases, the forecast showed acceptable performance right from the start of its operation (2 weeks after the start of data collection, where those 2 weeks are used to initialise the model). Detailed quantitative assessment of this performance is presented in this subsection.

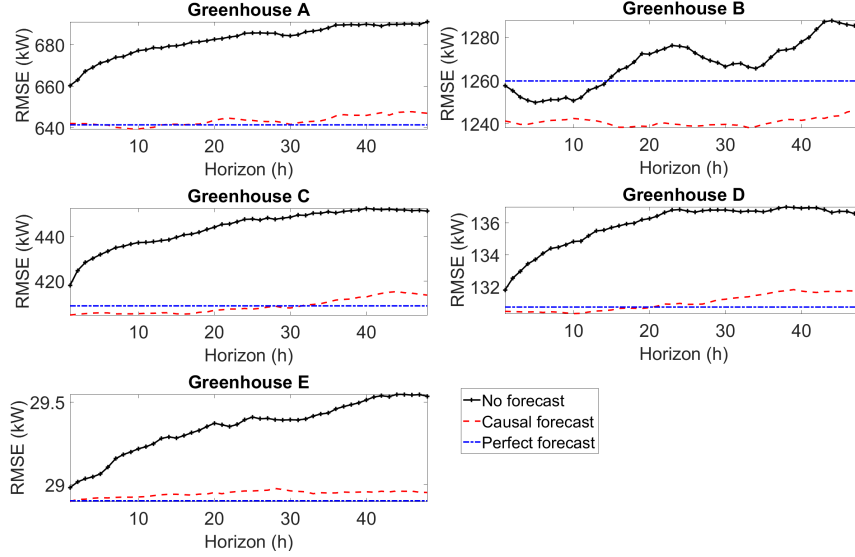


Figure 8: Comparison of performance between all 3 setups for each greenhouse (*causal forecast* setup).

Scatter plots of measured versus predicted heat load are presented in Fig. 9 (1 h ahead) and Fig. 10 (24 h ahead). One may notice a number of points along the bounds (either zero predicted and non-zero load, or 0 load measured and non-zero predicted) that translate unforeseen starts/stops of the heating supply.

The performance of the predictor fluctuates throughout time for all greenhouses, as changes occur in the operation of the greenhouses. To illustrate this using data from Greenhouse D, Fig. 11 shows a period where the predictor performs well while Fig. 12 presents a period where the predictor is less able to forecast very high frequency changes in the load conditions.

Comparison of the performance of the predictor among greenhouses is presented in Fig. 13 together with an assessment of performance improvement compared to a simple baseline predictor using the last 24 h load profile as the reference for future behaviour. RMSEs were normalised by the peak load over the period for the greenhouse in order to allow for comparison.

The naive approach estimating future load profile by a 24 h persistence of the load profile ("tomorrow = today") yielded an RMSE within 10–30% of the peak load. In comparison, the RMSE of the predictor of the proposed RLS predictor was within 8–20% of the peak load of the greenhouse. In 4 cases, the RMSE was reduced by more than 25%, and within 13–24% for the remaining greenhouse. It is worth noting that for both predictors the load of Greenhouse B was the least well predicted, followed by A, E, C and D.



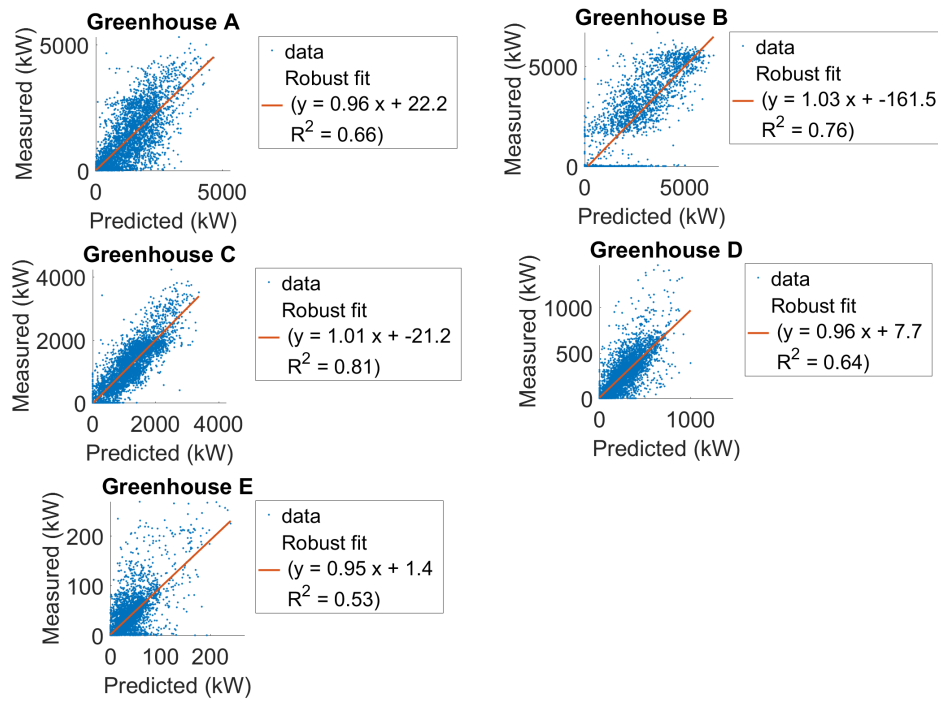


Figure 9: Predicted vs. actual heat load (1 h ahead - *causal forecast* setup).

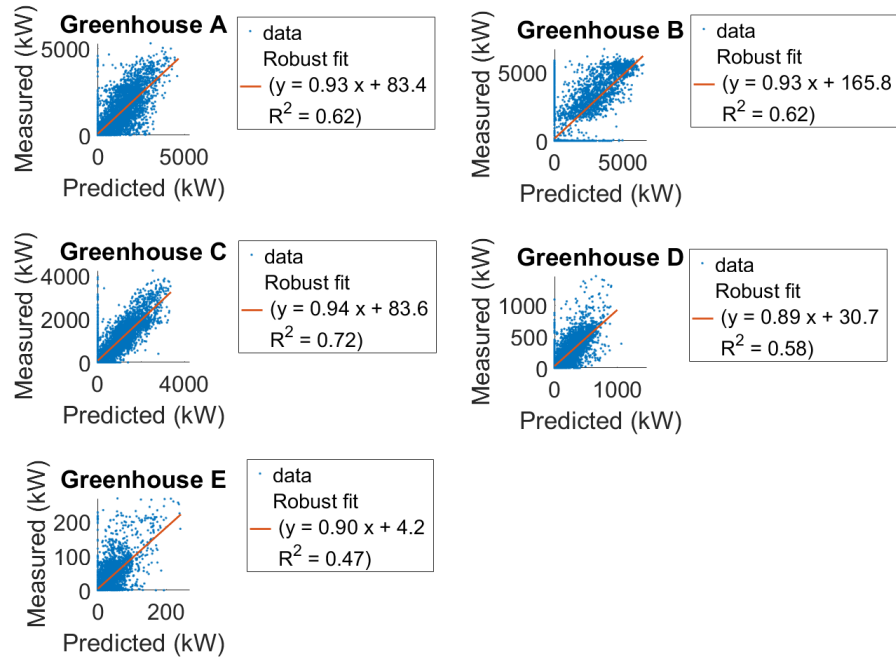


Figure 10: Predicted vs. actual heat load (24 h ahead - *causal forecast setup*).

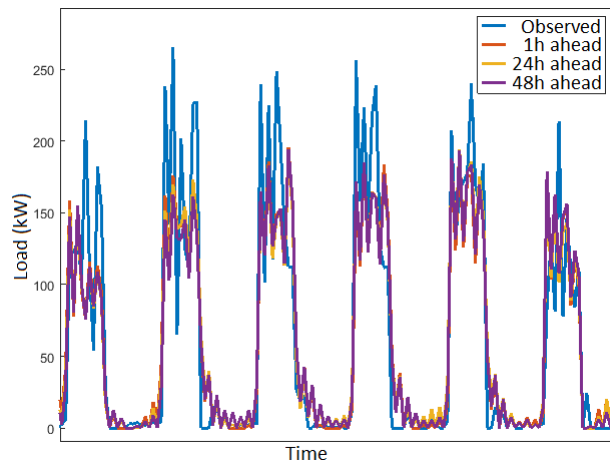


Figure 11: Greenhouse D: days where the prediction captures well the main load dynamics (*causal forecast setup*).

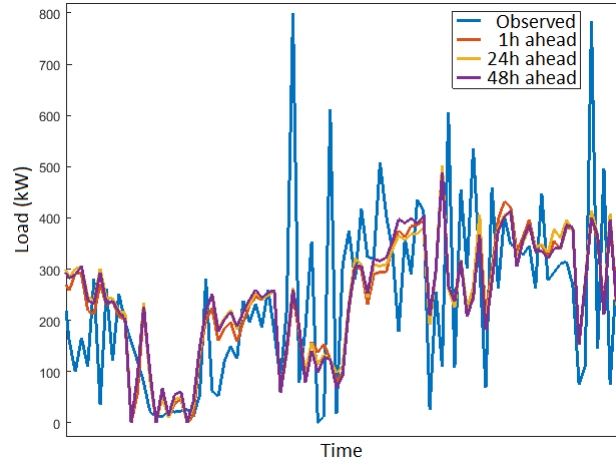


Figure 12: Greenhouse D: days where the prediction fails to capture load dynamics (*causal forecast setup*).

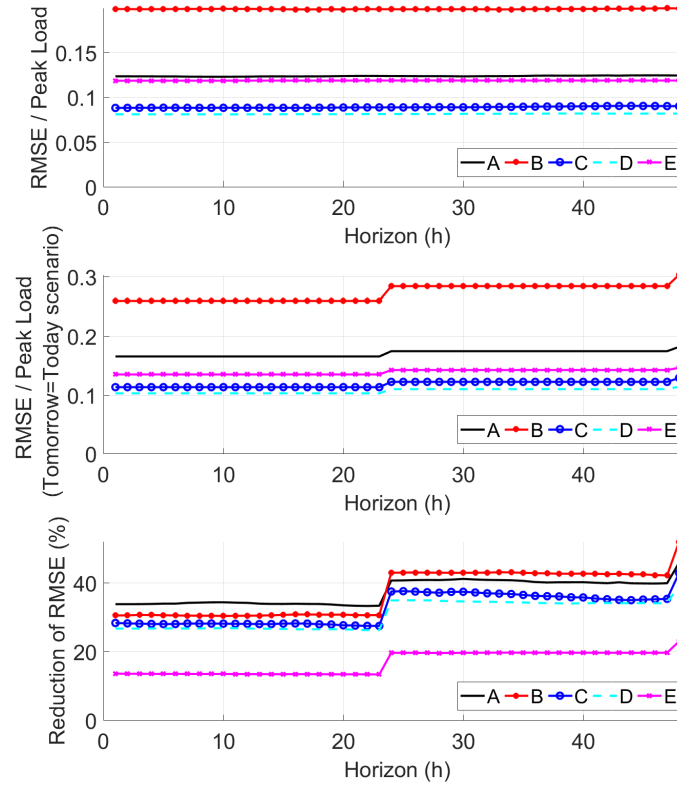


Figure 13: Comparison of prediction performance among greenhouses (*causal forecast setup*) against a naive "tomorrow=today" baseline.

### 3.4. Benefits from model extensions and tuning

Benefits from further model extensions and tuning were investigated. First, identification of a separate model structure for longer horizons (above 12 h) was investigated. Then, error modelling was investigated through the use of the latest residual as an extra explanatory variable.

In a third step, tuning was operated using the `fminsearchbnd` function in MATLAB, jointly optimising the value of the forgetting factor within the range  $[0.95, 1)$  (initialised at 0.99) and of the filter parameters (see Eq. 2) within the range  $[0, 0.5]$  (initialised at 0.1) to minimise the RMSE over the whole prediction period (a posteriori). This is only a rather rough optimisation, as the aim was to assess the relevance of the initial value selected for the parameters. However, the results suggest that the benefit of this tuning is minor (apart from Greenhouse B), therefore not justifying the investigation of more complex optimisation methods.

Lastly, combinations of these model extensions were also studied by considering separate identification and tuning of a short and long horizon model, using the latest residual as an extra explanatory variable. Resulting changes in performance are presented in Fig. 14. One may observe that, apart from Greenhouse B, all these extensions (as well as their combination) only had minor effects on the performance, as the RMSE varied by less than 2% of its value.

For Greenhouse B, tuning resulted in a lower value of the forgetting factor (0.97) which was found to bring a significant improvement of the RMSE (in the range of 10%). This may be explained by the resulting faster ability to adapt to the abrupt changes in this greenhouse's load. Moreover, although Fig. 7 suggested that an autoregressive component was missing in the model, the implementation of such an error model did actually lead to a counter-intuitive strong degradation of the RMSE for short horizons.

## 4. Discussion

Recursive least squares is a simple method for online adaptive forecast, which was observed to be effective for heat load prediction in greenhouses. Recurring patterns in the load were well captured by the predictor, as the weekly curve terms were among the most significant input predictors. Relevant weekly curve terms for the model denoted daily periodicity, but also slower (weekly) changes for most greenhouses. In every case, the weather forecast was also found to provide further improvement of the prediction, especially for long prediction horizons.

Fast frequency changes in the load conditions, such as spikes and fast random changes were often not well captured by the predictor. However, it is expected that such changes would average out when multiple greenhouses are aggregated in an area. Alternatively, if the focus is on a single greenhouse, one may consider the effect of communication by allowing the greenhouse owner to send a signal expressing a change in operating conditions (such as an interruption in operation, or a change in crop), or predicting the

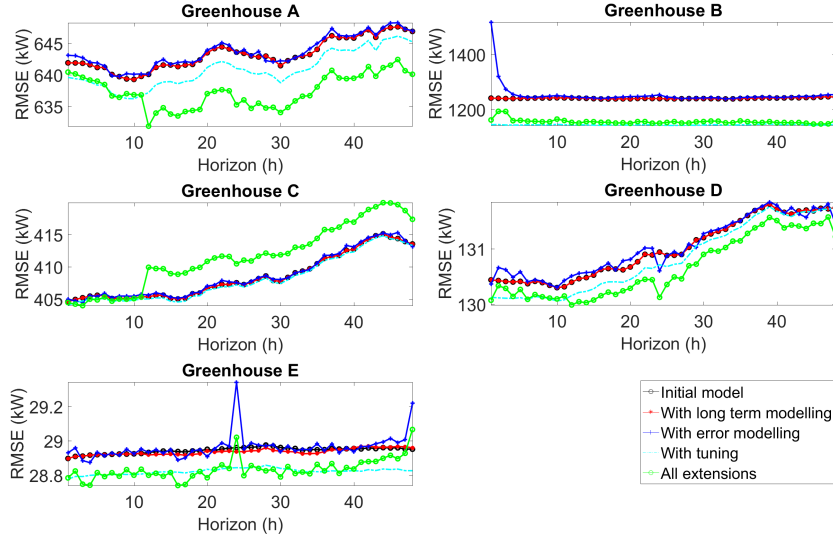


Figure 14: Benefits from model extension and additional tuning (*causal forecast setup*).

interruption. In such a setup, the use of a variable forgetting factor [40, chap. 11] could bring some improvement.

This investigation is however subject to a number of limitations. In particular, only a reduced number of greenhouses was analysed although they exhibited different characteristics. Moreover, the weather data originated from a weather station located remotely from the location (ca. 20 km), rather than local measurements.

Concerning the forward selection algorithm of relevant explanatory variables, a clear limitation is the fact that these were identified on the dataset a posteriori. In practise, this would require some history to identify those before using them in the recursive least squares estimator. The same remark holds for tuning of the forgetting factor and input filter parameters, although those parameters seemed to have little impact on the performance in 4 out of 5 cases.

However, in practise this could be alleviated through use of typical useful inputs initially (namely ambient temperature and weekly curve terms 0,7,14,21), values for forgetting factor (0.99) and filter coefficients (either 0.1 or none). This would allow temporarily running a sub-optimal estimator until sufficient data is collected to operate further identification and tuning.

## 5. Conclusion

In this study, the performance of a recursive least square forecast was evaluated for online short term greenhouse heat load prediction in a district heating system. Historical data from 5 greenhouses were used, together with weather measurements and a commercial weather forecast service.

Predictor input were automatically selected among weekly curve terms and available weather variables. Relevant inputs differed greenhouse by greenhouse, therefore supporting the use of individual models and such a forward selection approach. Weekly curve terms of a period higher than one day were identified as relevant in some cases, which suggests that the traditional approach consisting in the use of diurnal curves may miss relevant available information. A weather forecast was found to improve the performance of the prediction, although most of the variance reduction originated from the time dependency (modelled through weekly curves), and the quality of the forecast did affect which variables were relevant in two of the greenhouses. Tuning of the forgetting factor was found to provide improvement of performance in cases when the greenhouse had a number of fast significant changes in behaviour and many interruptions in operation.

This simple method led to a forecast with an root mean square error in the range of 8–20% of peak load for each of the 5 greenhouses over the period of the study.

Future work could consider a study on a larger number of greenhouses and their aggregated load, extending the study to a complete year of data to evaluate how this performance varies throughout time. It could also assess if greenhouses with a more automated heating control exhibit a more predictable heat load than those where the heating is mainly controlled by manual intervention. Additionally, once larger datasets covering years of operation become available, comparing the performance of the method used in this work with other types of predictors (with a particular focus on robustness) could yield interesting results.

## Acknowledgements

The authors would like to acknowledge support and valuable input from Lasse Elmelund Pedersen from *Fjernvarme Fyn A/S* and Per Dahlgaard Pedersen from *Neogrid Technologies ApS*, as well as provision of weather forecast data by Henrik Aalborg Nielsen from *ENFOR A/S*.

The research leading to these results is a collaboration between the ADVANTAGE project<sup>1</sup> (funded by the European Community's Seventh Framework Programme *FP7-PEOPLE-2013-ITN* under grant agreement no. 607774) and the CITIES project<sup>2</sup> (supported by the Danish Strategic Research Council).

## References

- [1] H. Lund, S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund, B. V. Mathiesen, 4th Generation District Heating (4GDH). Integrating smart thermal

---

<sup>1</sup><http://www.fp7-advantage.eu/>

<sup>2</sup><http://www.smart-cities-centre.org/>

- grids into future sustainable energy systems., *Energy* 68 (2014) 1–11. doi:10.1016/j.energy.2014.02.089.
- [2] H. Madsen, O. Palsson, K. Sejling, H. Søgaaard, Models and methods for optimization of district heating systems, part I: models and identification methods, Tech. rep., IMSOR, The Technical University of Denmark (1990).
  - [3] R. Campo, P. Ruiz, Adaptive Weather-Sensitive Short Term Load Forecast, *IEEE Transactions on Power Systems* 2 (3) (1987) 592–598. doi:10.1109/TPWRS.1987.4335174.
  - [4] T. Mestekemper, Energy Demand Forecasts and Dynamic Water Temperature Management, Ph.D. Thesis, Universität Bielefeld, 2011.
  - [5] M. Q. Raza, A. Khosravi, A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings, *Renewable and Sustainable Energy Reviews* 50 (2015) 1352–1372. doi:10.1016/j.rser.2015.04.065.
  - [6] B. Yildiz, J. Bilbao, A. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, *Renewable and Sustainable Energy Reviews* 73 (March 2016) (2017) 1104–1122. doi:10.1016/j.rser.2017.02.023.
  - [7] G. Sandou, S. Font, S. Tebbani, A. Hired, C. Mondon, Predictive Control of a Complex District Heating Network, in: *IEEE Conference on Decision and Control*, Seville, 2005, pp. 7372–7377.
  - [8] J. Eynard, S. Grieu, M. Polit, Optimal control of a multi-energy district boiler: A case study, in: *IFAC Proceedings Volumes (IFAC-PapersOnline)*, Vol. 18, IFAC, 2011, pp. 8271–8276. doi:10.3182/20110828-6-IT-1002.00139.
  - [9] H. Madsen, K. Sejling, H. T. Søgaaard, O. P. Palsson, On flow and supply temperature control in district heating systems, *Heat Recovery Systems & CHP* 14 (6) (1994) 613–620.
  - [10] F. Wernstedt, Multi-agent systems for distributed control of district heating systems, Ph.D. thesis, Blekinge Institute of Technology (2005).
  - [11] P. Pinson, T. S. Nielsen, H. A. Nielsen, N. K. Poulsen, H. Madsen, Temperature prediction at critical points in district heating systems, *European Journal of Operational Research* 194 (1) (2009) 163–176. doi:10.1016/j.ejor.2007.11.065.
  - [12] J. C. López, A. Baille, S. Bonachela, M. M. González-Real, J. Pérez-Parra, Predicting the energy consumption of heated plastic greenhouses in south-eastern Spain, *Spanish Journal of Agricultural Research* 4 (4) (2006) 289–296.
  - [13] O. Jolliet, L. Danloy, J. B. Gay, G. L. Munday, A. Reist, HORTICERN: an improved static model for predicting the energy consumption of a greenhouse, *Agricultural and Forest Meteorology* 55 (3-4) (1991) 265–294. doi:10.1016/0168-1923(91)90066-Y.

- [14] J. Chen, J. Yang, J. Zhao, F. Xu, Z. Shen, L. Zhang, Energy demand forecasting of the greenhouses using nonlinear models based on model optimized prediction method, *Neurocomputing* 174 (2016) 1087–1100. doi:10.1016/j.neucom.2015.09.105.
- [15] M. Trejo-Perea, G. Herrera-Ruiz, J. Rios-Moreno, R. Miranda, E. Rivas-Araiza, Greenhouse Energy Consumption Prediction using Neural Networks Models, *International Journal of Agriculture Biology* 11 (1) (2009) 1–6.
- [16] P. Bacher, H. Madsen, H. A. Nielsen, B. Perers, Short-term heat load forecasting for single family houses, *Energy and Buildings* 65 (2013) 101–112. doi:10.1016/j.enbuild.2013.04.022.
- [17] H. A. Nielsen, H. Madsen, Predicting the Heat Consumption in District Heating Systems using Meteorological Forecasts, Tech. Rep. December, DTU IMM (2000).
- [18] H. A. Nielsen, H. Madsen, Modelling the heat consumption in district heating systems using a grey-box approach, *Energy and Buildings* 38 (1) (2006) 63–71. doi:10.1016/j.enbuild.2005.05.002.
- [19] S. Idowu, S. Saguna, C. Åhlund, O. Schelén, Forecasting Heat Load for Smart District Heating Systems: A Machine Learning Approach, in: *IEEE International Conference on Smart Grid Communications Forecasting*, 2014, pp. 554–559.
- [20] M. Dahl, A. Brun, G. B. Andresen, Using ensemble weather predictions in district heating operation and load forecasting, *Applied Energy* 193 (2017) 455–465. doi:10.1016/j.apenergy.2017.02.066.
- [21] S. Grosswindhager, A. Voigt, M. Kozek, Online short-term forecast of system heat load in district heating networks, In *Proceedings of the 31st International Symposium on Forecasting* (1) (2011) 1–8.
- [22] M. Protić, S. Shamshirband, D. Petković, A. Abbasi, M. L. Mat Kiah, J. A. Unar, L. Živković, M. Raos, Forecasting of consumers heat load in district heating systems using the support vector machine with a discrete wavelet transform algorithm, *Energy* 87 (2015) 343–351. doi:10.1016/j.energy.2015.04.109.
- [23] H. H. Nguyen and C. W. Chan, “Multiple neural networks for a long term time series forecast,” *Neural Computing and Applications*, vol. 13, no. 1, pp. 90–98, 2004.
- [24] R. Jovanović, A. A. Sretenović, B. D. Živković, Ensemble of various neural networks for prediction of heating energy consumption, *Energy and Buildings* 94 (2015) 189–199. doi:10.1016/j.enbuild.2015.02.052.
- [25] H. Hippert, C. Pedreira, R. Souza, Neural networks for short-term load forecasting: a review and evaluation, *IEEE Transactions on Power Systems* 16 (1) (2001) 44–55. doi:10.1109/59.910780.



- [26] S. Shamshirband, D. Petković, R. Enayatifar, A. Hanan Abdullah, D. Marković, M. Lee, R. Ahmad, Heat load prediction in district heating systems with adaptive neuro-fuzzy method, *Renewable and Sustainable Energy Reviews* 48 (2015) 760–767. doi:10.1016/j.rser.2015.04.020.
- [27] H. H. Çevik and M. Çunkaş, “Short-term load forecasting using fuzzy logic and ANFIS,” *Neural Computing and Applications*, vol. 26, no. 6, pp. 1355–1367, 2015.
- [28] M. Protić, S. Shamshirband, M. H. Anisi, D. Petković, D. Mitić, M. Raos, M. Arif, K. A. Alam, Appraisal of soft computing methods for short term consumers’ heat load prediction in district heating systems, *Energy* 82 (2015) 697–704. doi:10.1016/j.energy.2015.01.079.
- [29] J. Eynard, S. Grieu, M. Polit, Wavelet-based multi-resolution analysis and artificial neural networks for forecasting temperature and thermal power consumption, *Engineering Applications of Artificial Intelligence* 24 (3) (2011) 501–516. doi:10.1016/j.engappai.2010.09.003.
- [30] M. J. Sanjari, H. B. Gooi, Probabilistic Forecast of PV Power Generation Based on Higher-Order Markov Chain, *IEEE Transactions on Power Systems* 32 (4) (2017) 1–1. doi:10.1109/TPWRS.2016.2616902.
- [31] J. Nowotarski, B. Liu, R. Weron, T. Hong, Improving short term load forecast accuracy via combining sister forecasts, *Energy* 98 (2016) 40–49. doi:10.1016/j.energy.2015.12.142.
- [32] H. X. Zhao, F. Magoulès, A review on the prediction of building energy consumption, *Renewable and Sustainable Energy Reviews* 16 (6) (2012) 3586–3592. doi:10.1016/j.rser.2012.02.049.
- [33] D. Petković, M. Protić, S. Shamshirband, S. Akib, M. Raos, D. Marković, Evaluation of the most influential parameters of heat load in district heating systems, *Energy and Buildings* 104 (2015) 264–274. doi:10.1016/j.enbuild.2015.06.074.
- [34] J. W. Taylor, R. Buizza, Neural Network Load Forecasting With Weather Ensemble Predictions, *IEEE Transactions on Power Systems* 17 (3) (2002) 626–632.
- [35] L. Arvastson, Stochastic Modeling and Operational Optimization in District Heating Systems, Ph.D. thesis, Lund University (2001).
- [36] E. Dotzauer, Simple model for prediction of loads in district-heating systems, *Applied Energy* 73 (3-4) (2002) 277–284. doi:10.1016/S0306-2619(02)00078-8.
- [37] G. H. John, R. Kohavi, K. Pfleger, Irrelevant Features and the Subset Selection Problem, *Machine Learning Proceedings 1994* (1994) 121–129doi:10.1016/B978-1-55860-335-6.50023-4.

- [38] A. Poulin, M. Dostie, M. Fournier, S. Sansregret, Load duration curve: A tool for technico-economic analysis of energy solutions, *Energy and Buildings* 40 (1) (2008) 29–35. doi:10.1016/j.enbuild.2007.01.020.
- [39] K. Difs, M. Danestig, L. Trygg, Increased use of district heating in industrial processes - Impacts on heat load duration, *Applied Energy* 86 (11) (2009) 2327–2334. doi:10.1016/j.apenergy.2009.03.011.
- [40] H. Madsen, *Time series analysis*, Chapman & Hall, 2008.
- [41] P. J. C. Vogler-finck, P. Bacher, Predicting the heat load of greenhouses in a district heating system - A case study in Fyn, Tech. rep., Neogrid Technologies ApS, DTU Compute (2016).